DOCUMENT RESUME

ED 287 290 FL 016 946

AUTHOR McIntyre, Sandra
TITLE Investigations into the Appropriate Use of Various Item Types for Testing Specific Levels of Reading Proficiency.
PUB DATE 87
NOTE 23p.; In: Language Testing Research; Selected Papers from the Colloquium (Monterey, California, February 27-28, 1986); see FL 016 938.
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Cloze Procedure; *Difficulty Level; French; Item Analysis; Korean; Language Proficiency; *Language Tests; Mandarin Chinese; Rating Scales; Reading Comprehension; Reading Skills; *Reading Tests; Second Language Instruction; Spanish; *Test Format; *Test Items; Uncommonly Taught Languages
IDENTIFIERS *Defense Language Institute; *Interagency Language Roundtable

ABSTRACT

A study investigated the appropriate use of item types for testing reading at various levels of proficiency. It was specifically concerned with item types found in the standardized group-administered Defense Language Proficiency Tests. The item formats investigated were signs in the target language, identification of underlined information, English language comprehension questions following a target language passage, true-false-not addressed questions, and cloze procedure. Responses to specific item types on currently-used tests of Spanish, Korean, Mandarin Chinese, and French were analyzed and examinee comments on the items were considered. Extensive use of the Interagency Language Roundtable's oral interview system was made in the analysis. Appended materials include data on item-type responses on the four language tests and examples of the item formats. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED287290

INVESTIGATIONS INTO THE APPROPRIATE USE OF VARIOUS ITEM TYPES FOR TESTING SPECIFIC LEVELS OF READING PROFICIENCY

Sandra McIntyre

Defense Language Institute

The purpose of this paper is to report results of preliminary investigations carried out by the Defense Language Institute concerning the appropriate use of item types for testing reading at various levels of proficiency. The investigations made extensive use of the language proficiency scale associated with the Interagency Language Roundtable (ILR) oral interview system and were specifically concerned with item types found in the standardized group-administered Defense Language Proficiency Tests.

Background

The ILR Oral Interview and the DLPT III

The language proficiency scale of the Interagency Language Roundtable is an eleven-point scale which defines levels of language proficiency for each of four skills: Speaking, Listening, Reading, and Writing.

The scale originates from the Interagency Language Roundtable Oral Interview system, which has been used by various government agencies as an official system for evaluating language proficiency for more than 30 years. During that time the proficiency levels have been carefully defined and they now serve as the basis for group-administered tests as well as for the face-to-face oral interviews. Some of the most widely used group-administered language proficiency tests are the Defense Language Proficiency Tests (DLPTs).

The DLPT III Reading Test

The Lower Range DLPT III Reading Test is a group-administered paper-and-pencil test which is machine scorable. It has five to seven parts, 100 to 120 items, and takes three hours to administer. Examinees are not permitted to use lexical aids or take notes. Item types that may be found in the DLPT III include the identification of signs, the identification of underlined information, factual and inferential comprehension questions, true-false-not addressed questions, and a modified version of the cloze.

The DLPT IIIs are based on the proficiency scale developed, refined and approved by the Interagency Language Roundtable. They are designed, written and validated according to the Interagency Language Roundtable Skill Level Descriptions. The validation process yields information regarding the correlation between proficiency levels of the examinees, as measured by the face-to-face assessment of reading during the oral interview, and the total scores of the examinees on the reading portion of the DLPT III. The validation data also provide an opportunity to examine the relationship between the proficiency levels of the examinees and the performance of the examinees on each part of the test.

FL016946

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R Schoffman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

93

2

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Each part of the test uses an item format intended to measure a specific level of proficiency. Therefore, by examining the relationship between the proficiency levels of the examinees (as defined by the ILR) and their performance on each part of the test, information can be gathered concerning the appropriateness of using the item formats for testing specific levels of proficiency. The relationship can be examined by graphing the performance of the examinees at each level of proficiency.

For example, Appendix A shows the performance of examinees on the Reading Test of the Spanish DLPT III. The reading proficiency scores of the examinees, as determined by the reading portion of the face-to-face Interagency Oral Proficiency Interview, are shown along the x-axis. The percent of the items answered correctly on the DLPT III reading test is shown on the y-axis. Each of the curves represents a part of the test. For example, Part I, which consists of the "sign" format is represented by the curve at the top of the graph.

The intercept and the relative position of the curve in relationship to the other curves provides information concerning the difficulty of the subtest. In this case, the curve for Part I seems to indicate that it is easier than the other parts. Examinees at each of the six levels (0+ through 3) answered a higher percent of the items on Part I correctly than they did on the other parts. The shape of the curve provides information concerning the ability of the subtest to discriminate between the examinees of various levels of proficiency. The curve for Part I shows a sharp incline between the level 0+ and level 1 examinees. On the average, level 0+ examinees responded correctly to 50% of Part I items, whereas level 1+ examinees responded correctly to 75% of Part I items. The slope suggests that the item format in Part I discriminates well between examinees at levels 0+ to 1. Thus, by examining the curves for each of the item formats used in several DLPT IIIs for different languages, information has been gathered concerning the appropriate use of specific item formats for testing at the various levels of proficiency.

Formats

Signs

Part I of the DLPT III uses signs in the language being tested followed by four options in English. For example:

Exit (written in the target Language)

This sign says:

A. stop
B. exit
C. slow
D. caution

The sign format used in Part I was designed to test at the lowest levels of proficiency. The data graphed on curves generally support the intended use of the format. As previously mentioned, the results from the Spanish DLPT III (see

Appendix A) demonstrate that Part I was the easiest part for the examinees and that it discriminated best between levels 0+ and 1. The data obtained for the Korean, Chinese and French DLPT IIIs show a similar pattern of results for the sign format. (See Appendices B, C and D.) In each case Part I is relatively easy in comparison to the sections of the test which were designed to test beyond level 1+. In addition, the slopes for Part I indicate that the format served to discriminate between examinees at levels 0+ to 1. Surprisingly, the curve of the Chinese DLPT III indicates a rise in the slope between the Part I scores of level 1+ and level 2 examinees which is larger than the rise between the level 0+ and 1 examinees. This slight aberration may be attributed to the small N of the 0+ group of students. It may also indicate the particular nature of Chinese signs which tend to represent relatively condensed language and low frequency character patterns. The overall results of the sign format, however, indicate that the format seems to be working effectively to discriminate among examinees at the lower end of the proficiency scale.

Information Identification

Appendix E provides an example of a second item type: the identification of underlined information. In the DLPT III the paragraph is in the language of the test and the options are in English. By varying the difficulty level of the paragraph, the underlined items and the distractors, the overall difficulty level of the format can be altered. The format has been used in the DLPT III to test reading proficiency at levels 1+ and 2+. Appendix F is an example of information identification a level 2/2+.

Both the Korean and the Chinese DLPT IIIs contain information identification items at levels 1+ and 2+ (see Appendices B and C). The relative positions of the curves suggest that the part containing the level 2+ items is more difficult than the part containing the level 1+ items. In both the Korean and the Chinese DLPT IIIs the level 2+ section continued to discriminate up through level 3, whereas the level 1+ section failed to discriminate between examinees at level 2+ and 3.

In summary, the information identification format has been used at levels 1+ and 2+. Results suggest that some formats can be used for testing more than one level by changing the difficulty of both the stimulus and the options.

Comprehension Questions

Another format which has been used at different levels consists of a passage, in the target language, followed by comprehension questions in English (see Appendix G). The DLPT III reading tests have used this format with level 1+ passages and with level 2+ to 3 passages. For the level 2+/3 section of the test, the difficulty of both the passages and the questions is increased. The questions used for testing at level 1+ ask for concrete factual information. Those used for testing at level 2+/3 tend to be much more inferential in nature. Although developing multiple choice comprehension questions at the 2+/3 level requires more effort than developing questions at the 1+ level, the format does seem to work as well at the higher level as at the lower level.

Part II of the Korean DLPT III (see Appendix B) consists of short passages, such as common announcements, followed by factual questions. The difficulty level of Part II is comparable to that of Part I. The curve for Part II also tends to follow approximately the same slope as Part I, though Part II discriminates more powerfully than Part I between levels 1 and 1 + .

Part IV of the Chinese DLPT III (see Appendix C) consists of relatively long passages of a more abstract, theoretical nature, followed by inferential multiple choice questions. The relative placement of the curve suggests that the examinees found it difficult. In addition, Part IV discriminates well among the higher level examinees. In conclusion, the multiple choice comprehension questions format has been used successfully to test higher levels of proficiency (2 + /3) as well as the lower levels of proficiency.

True-False-Not Addressed

Another format which has been used to test reading at the abstract/inferential levels (levels 3 through 3 +) is the True-False-Not Addressed format. An example can be found in Appendix H.

The DLPT III contains authentic passages in the target language followed by statements in English. The examinees are instructed to determine, on the basis of the information given in the passage, whether a statement is true, false, or not addressed. A statement is true if it can be confirmed on the basis of the information given in the passage. It is false if it is contradicted by information given in the passage, and it is not addressed if there is insufficient information in the passage to either confirm or deny the statement. Although this format may appear to be fairly easy to construct, careful attention must be given to determining and confirming through field trial that each statement is, in fact, clearly true, false, or not addressed. This format, however, has been found to work well in testing the ability to understand ideas and inferences at level 3/3 + .

Part V of the Chinese DLPT III reading test is more difficult than Part IV for the students who are at the level 2 to 3 range of proficiency. (See Appendix C.) The reversal of the curves for level 0 + and 1 + examinees may be attributed to the fact that the chance of correctly answering an item in Part V is higher than the chance of correctly answering an item in Part IV. (Part V has only three options from which to select, whereas Part IV has four options.) The steep and steady incline in the slope for the Part V curve among the level 1 + through level 3 examinees illustrates that the True-False-Not Addressed format is useful at this level.

Cloze

The traditional cloze format, in which every seventh word is randomly chosen for deletion from the text and the examinees are asked to use their intuition to predict the best word to fill in the blank, has been significantly modified for use in the DLPT III. Because the DLPT III is a machine-scorable test, a list of options must be supplied. In addition, the lower level cloze passages are prefaced by an English paraphrase of the passage. Appendix I is an example of the format.

The results of the cloze format in the Spanish DLPT III (see Appendix A) are fairly representative of the results found for other languages as well. In general, the cloze passages tend to be quite difficult. The relative positions of the curves show a large increase in difficulty level between Part I (level 0 + and 1) and Part 2 (level 2 cloze), and the Part IV underlined information (level 2 +) is easier than the level 2 cloze.

Verbal comments given by the examinees also indicated that the cloze format was generating a high level of frustration. Much of the difficulty may have been due to a test-administration limitation in which the examinees were not permitted to mark in the booklets. Therefore, there was an additional load on short-term memory from the requirement of having to remember which options had already been selected.

The cloze format, however, does seem almost unique in its ability to discriminate with a very high reliability and power across a wide range of reading proficiency levels. Each of the curves for the cloze format shows a consistent, steep incline in the slope from level 1 through level 3. Statistically, the cloze works well, and it is relatively easy to construct in comparison to the multiple choice and true-false-not addressed formats.

It has been suggested that further investigations into the difficulty level of the cloze format might include using the same passages which were used for the cloze in an alternate format (such as true-false-not addressed, multiple choice comprehension questions, or traditional cloze formats). Such investigations would help to pull apart the interaction between the effects of item format and difficulty level of the passage.

Conclusion

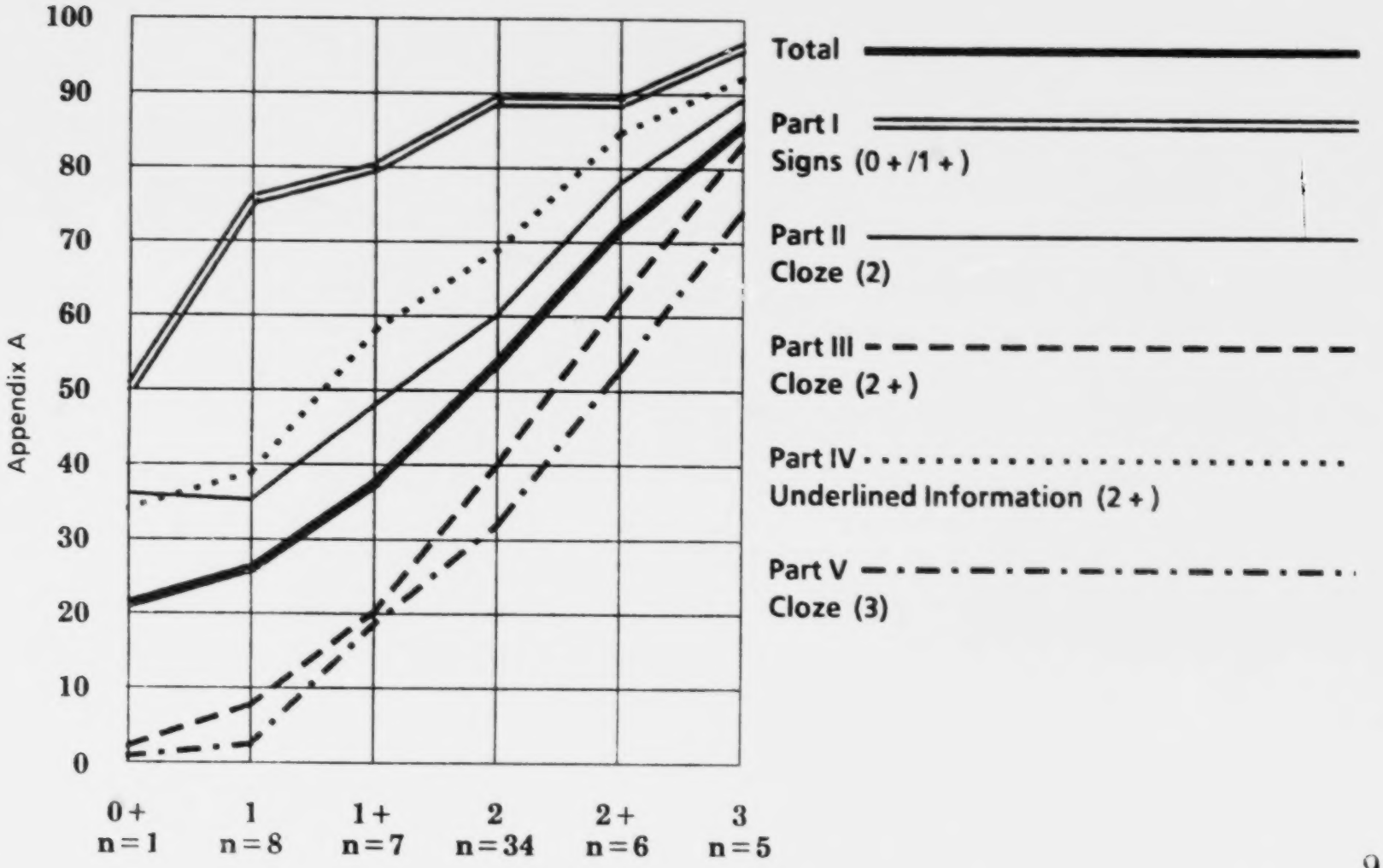
At DLI we have incorporated the results of these investigations to make the DLPT III reading tests even more effective as a measure of reading proficiency. The first DLPT IIIs, for example, employed the design exemplified by the Spanish test, and although the test had high reliability and validity, it generated complaints from examinees. As illustrated in the graph, the jump in difficulty between Part I and Part II was quite large. The general frustration level led examinees to question the face validity and suitability of the task. As a result, a new design was developed for the DLPT III reading test.

The new design incorporated several changes. The information format was moved to an earlier part of the test, and the cloze was replaced by inferential comprehension questions and true-false-not addressed items. Both of these formats had previously been used successfully in the listening portion of the DLPT III. The progression of the difficulty levels for the parts of the new DLPT III reading design has been satisfactory, and the reliability coefficients have remained strong.

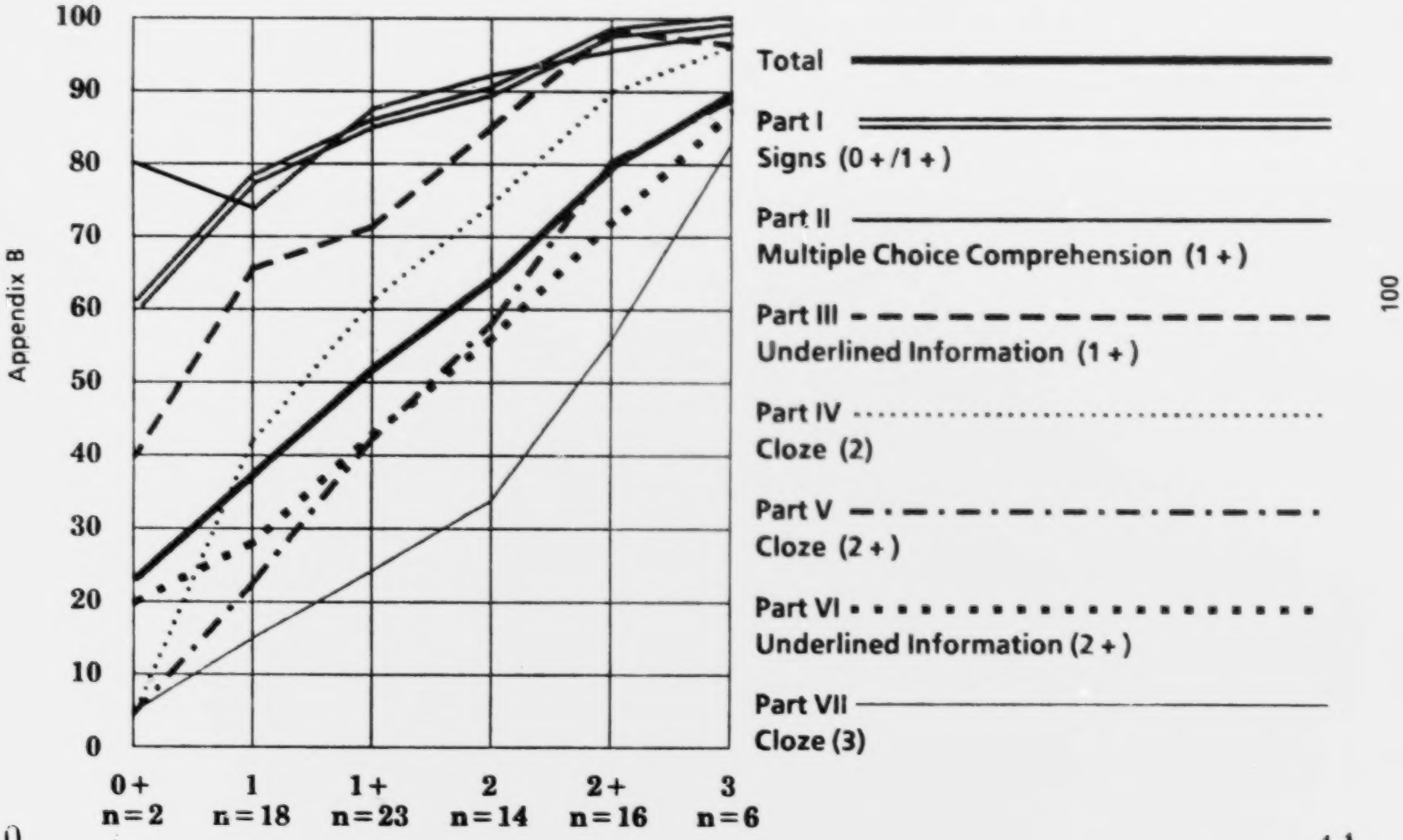
In conclusion, it should be noted that the ILR scale has been very useful not only in the initial development stages of the DLPT III, but also in its later validation and revision. The ILR scale has played a major role in optimizing the validity of the DLPT III tests. Not only are the test designs based upon the ILR definition, which give them a certain inherent face validity, but the results of the test can also be examined by using the ILR scale to determine criterion-related

validity. The oral interview, which can be used to assess three skills face-to-face, serves as an external criterion with which a correlation can be examined between total performance on the DLPT III and on the Oral Interview. In addition, the parts of the test can be examined based on Oral Interview performance. Without an integrated system based on a well defined scale, it would be a more difficult task to interpret the results of similar statistical analyses. Basing a test on a well-defined scale facilitates the design, development and validation of the test.

Spanish DLPT III
Reading



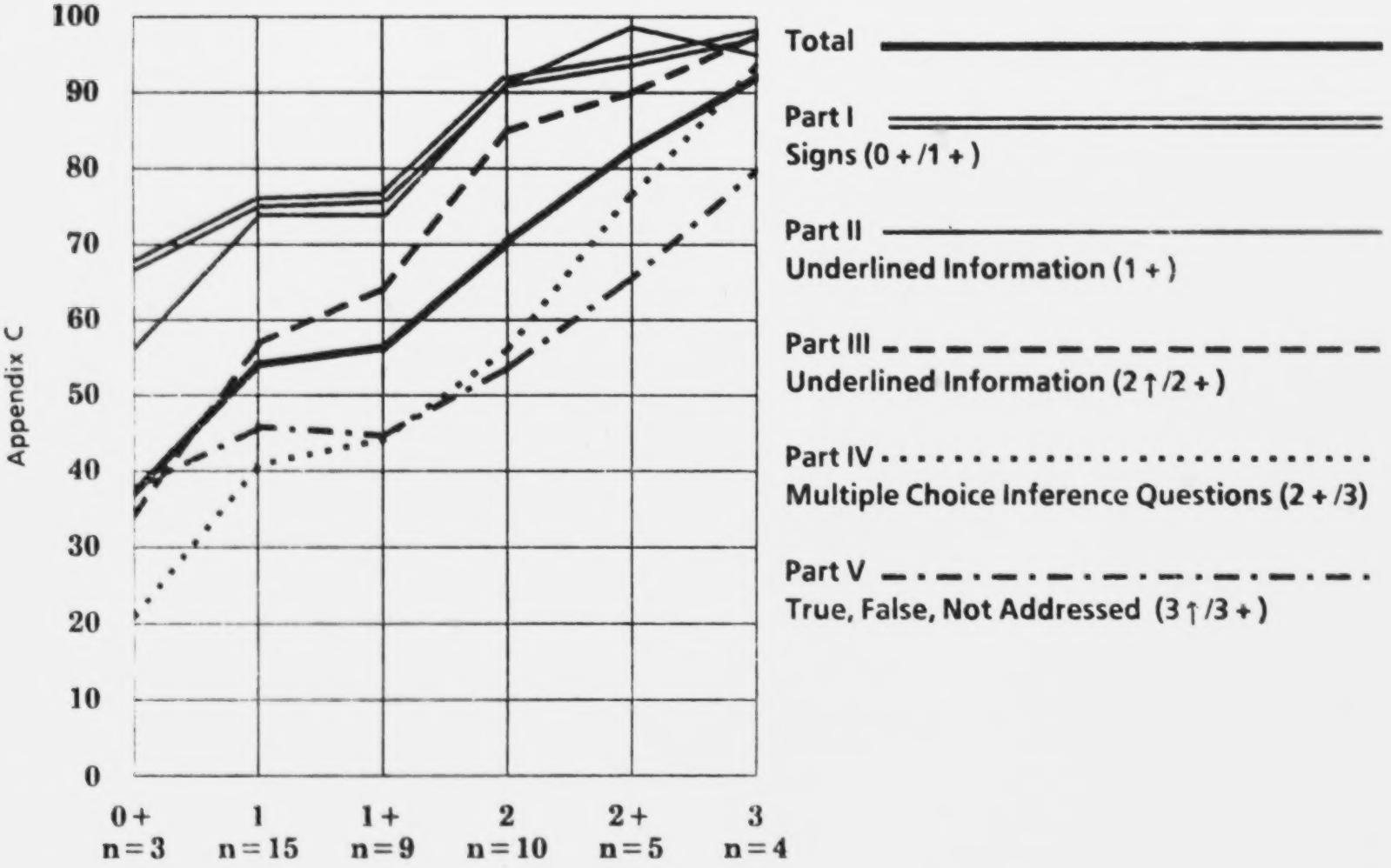
Korean DLPT III
Reading



Appendix B

100

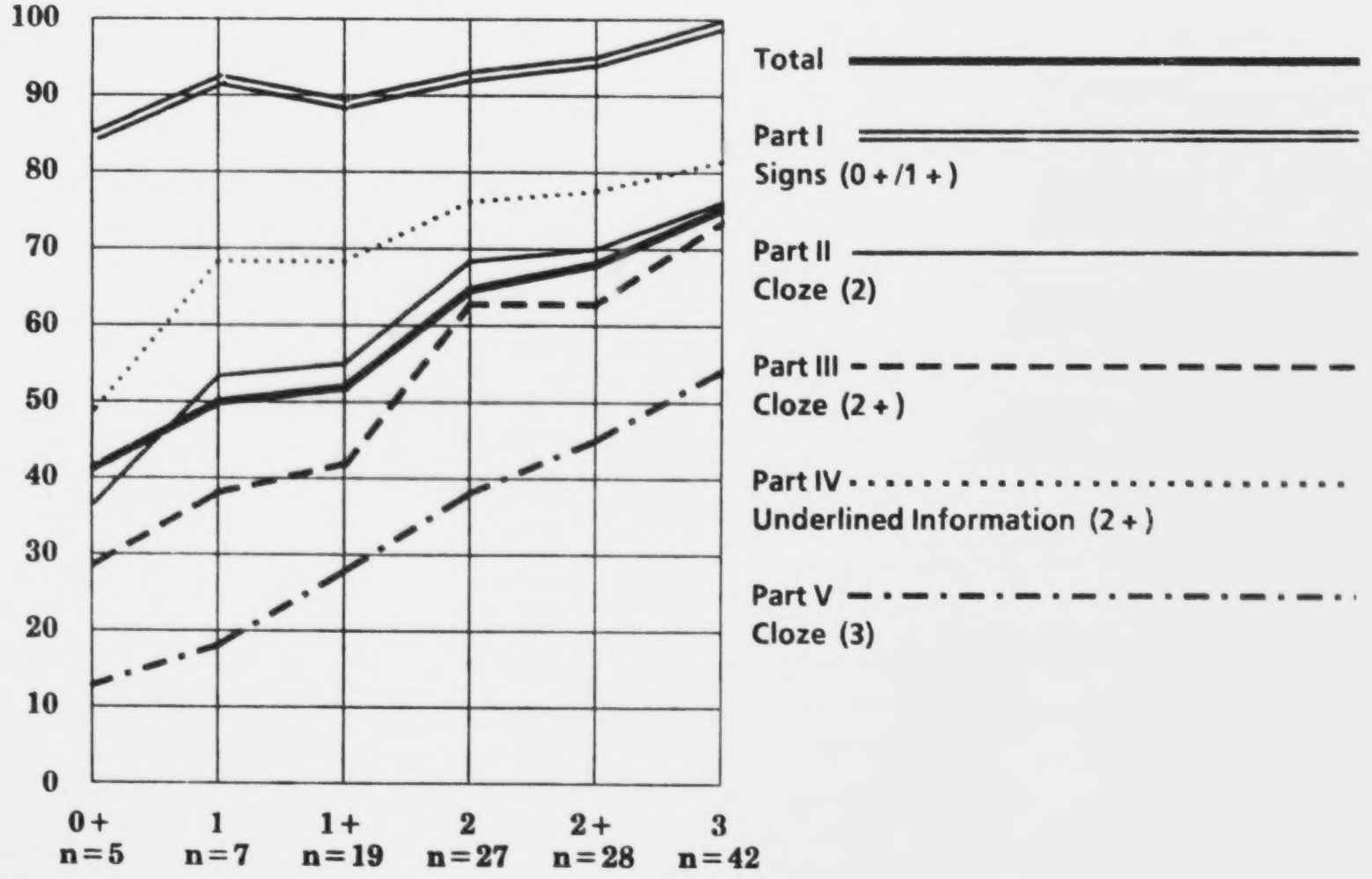
Chinese Mandarin DLPT III
Reading



Appendix C

101

French DLPT III
Reading



Appendix D

102

Appendix E

Information Identification, Level 1 +

Human Relations Meeting Set

The Seaside Human Relations Commission will hold a special meeting (1) today to plan activities (2) and projects for the year (3). The meeting will be held (4) at 5:30 p.m. in the City Hall (5) Conference room.

1.
A. meeting
B. dance
C. dinner
D. show

2.
A. to plan activities
B. planned to be active
C. plan to act it out
D. had planned to act on it

3.
A. for the year
B. after one year
C. a year ago
D. took a year

4.
A. will be held
B. was held
C. will be given
D. was given

5.
A. City Hall
B. Community Center
C. Concert Hall
D. Public Library

Appendix F

Information Identification, Level 2/2 +

Police Arrest Reeling Horseman For Drunken Driving After Chase

FREEMONT (AP)--A man who rode his horse home from a bar was arrested for (1) investigation of drunken driving (2) after a policeman saw him nearly fall out of the saddle (3).

John Charles Black was released on $1,500 bail (4) after his arrest early Monday. He was also accused of resisting arrest (5) because the horse allegedly (6) galloped off when the pursuing (7) police officer turned on his siren (8).

His wife, Tammi, said she and her husband often ride home on horseback from bars. "After all, you can't hurt yourself (9) on a horse and the horse knows its way back," she said. "It's the way to go if you're going to go out drinking.

"The horse watches out for cars ... It gives you time to sober up (10) between the bar and home."

But Freemont police said a horse falls under (11) the state Vehicle Code (12) definition (13) of a vehicle: anything that can be "propelled (14), moved or drawn upon a highway (15)."

1.

A. arrested for
B. questioned about
C. wanted for
D. suspected of

2.

A. drunken driving
B. reckless driving
C. improper conduct
D. speeding

3.

A. fall out of the saddle
B. collapse in his car
C. fall into this sad state
D. speed into the divider

4.

A. released on $1,500 bail
B. paid a fine of $1,500
C. held on a $1,500 bail
D. paid $1,500 in damages

5.

A. accused of resisting arrest
B. requested to stay still
C. warned to slow down
D. wanted to take a rest

6.

A. allegedly
B. quickly
C. willingly
D. suddenly

7.

A. pursuing
B. suspecting
C. determined
D. unsuspecting

8.

A. turned on his siren
B. turned up in his path
C. turned in his license
D. turned up his bullhorn

9.

A. hurt yourself
B. hurt others
C. be hurt by others
D. be hurting others

10.
A. sober up
B. unwind
C. exercise
D. slow down

11.
A. falls under
B. falls down
C. slips below
D. slips up

12.
A. Vehicle Code
B. Driver's License
C. Code of Conduct
D. Driver's Manual

13.
A. definition
B. law
C. expiration
D. explanation

14.
A. propelled
B. flown
C. sped
D. ridden

15.
A. highway
B. runway
C. street
D. alley

Appendix G

Comprehension Questions, Level 2 + /3

A Soviet scheme to divert the flow of two great Siberian revers southward to irrigate new agricultural lands "may have significant repercussions for the weather patterns of at least a hemisphere, if not the whole globe," writes physicist John Gibbin in the liberal Guardian of London (July 16).

Soviet planners are determined to reverse much of the flow of the Pechora River and the Ynisei and Ob Rivers. This will provide vital irrigation to millions of acres in Kazakhstan, a region opened to agriculture only in the past thirty years, and triple the grain production in the huge area. Dr. Gibbon says, "as far as it is possible to tell, the Soviet planners have dismissed the possibility that the side effects might do more harm than the irrigation does good."

The diversion will reduce by 20 percent the fresh water flow into the Arctic Ocean, "The ice-covered Arctic is the key factor in establishing the climatic patterns of the Northern Hemisphere." The diversion will allow the warmer salt water to rise to the surface and melt the ice. "By reducing the fresh water flow and causing the ice to break up, the irrigation schemes may directly reduce the rainfall in exactly the regions they are designed to help."

1. According to the article, the two great Siberian rivers

   A. are a matter of future international concern.
   B. are gradually yielding less water for irrigation.
   C. are threatened by changes in global weather patterns.
   D. currently flow southward into the grain producing region.

2. The plan will

   A. increase the water supply from the source into the rivers.
   B. revitalize ancient agricultural areas.
   C. increase the production in a relatively new agricultural region.
   D. forestall the effects of the expected weather pattern changes.

3. Dr. Gibbon feels that

   A. the Soviets have not recognized the possible repercussions.
   B. the plan will help the Soviets at the expense of other nations.
   C. the Soviets are wrong in expecting major weather pattern changes.
   D. the benefits of the plan outweigh the possible side effects.

4. Twenty percent of the fresh water

   A. will no longer flow into the Arctic Ocean.
   B. will come from melted ice from the Arctic Ocean.
   C. is expected to be lost into the ocean in the next few years.
   D. is expected to evaporate during the coming year.

5. According to the article

   A. the Soviets' plan may backfire.
   B. the plan is admirable and demonstrates foresight.
   C. weather patterns can't be controlled by man.
   D. agriculture is at the mercy of natural weather conditions.

Appendix H

True-False-Not Addressed, Level 3

BELGRADE, Yugoslavia--The Yugoslav leadership has warned the press to follow the party line and resist the arguments of opponents of the country's unorthodox self-management system of communism.

A statement by the nine-man collective presidency last week said "forces of international reaction" are bringing pressure on Yugoslavia in an effort to destabilize it.

Their object is "to bring in doubt and spread despondency and disbelief in Yugoslavia's capacity to overcome its problems," the statement maintained.

Yugoslavia's grave economic problems include debts to the West of $20 billion, high inflation, an acute shortage of hard currency, stagnating industrial production and falling living standards.

The statement said hostile forces, which it did not identify, are trying to penetrate Communist institutions and to influence the media. Media editorial boards are giving a distorted picture of Yugoslav reality, it said.

The statement did not specify which articles are in disfavor.

Questions:

1. The Yugoslav leadership supports an unorthodox system of Communism. (T)

2. The leadership attributes the unrest to international factors rather than to internal economic problems. (T)

3. The reactionary forces claim that a crisis is inevitable if the $20 billion national debt continues to grow. (NA)

4. The information disseminated by the media continues to reflect the ideology of the leadership. (F)

5. The leadership warned that authors of reactionary articles face the threat of arrest and imprisonment. (NA)

Appendix I

Paraphrase

Cloze, Level 2

On Thursday a 21 year old man identified as Michael W. Phillips from Seaside was hurt in an accident involving a motorcycle and a car. The driver of the car was Ho Jong Lee, 45, of Milpitas.

The accident occured in Monterey shortly before 3 p.m. Phillips will be treated before being released from Community Hospital.

Passage

Motorcyclist Injured In Collision

A seaside man was ____(★)____ Thursday ____(●)____ when ____(●)____
1 2 3
motorcycle ____(★)____ with a car on Garden Road in Monterey.
4

____(●)____ identified the injured ____(●)____ as Michael W. Phillips, 21
5 6
of 1148 Buena Street.

He ____(■)____ injured in ____(★)____ with a ____(●)____ driven ____(■)____
7 8 9 10
Ha Jong Lee, 45, of Milpitas, according to ____(■)____ police ____(★)____.
11 12

Phillips was ____(★)____ to Community hospital, ____(●)____ a spokesman
13 14
____(★)____ it was expected ____(■)____ he would ____(■)____ released ____(●)____
15 16 17 18
treatment.

The accident, which occurred ____(●)____ before 3 p.m., was ____(■)____
19 20
Garden Road north of Olmstead Road.

★

A. reported
B. report
C. taken
D. injured
E. collided
F. collision
G. said

●

A. where
B. when
C. after
D. just
E. afternoon
F. policeman
G. police
H. car
I. motorcyclist
J. his

■

A. was
B. be
C. on
D. by
E. a
F. that